

04-29-00

A

Please type a plus sign (+) inside this box → [+]

PTO/SB/05 (12/97)

<b>UTILITY PATENT APPLICATION TRANSMITTAL</b>  (Only for new nonprovisional applications under 37 CFR 1.53(b))	Attorney Docket No.	WILSONLESSAC	Total Pages	40
	First Named Inventor or Application Identifier			
	WILSON et. al.			
	Express Mail Label No.	EL465269058US		

<b>APPLICATION ELEMENTS</b>  See MPEP chapter 600 concerning utility patent application contents	TO <b>Assistant Commissioner for Patents Box Patent Application Washington, DC 20231</b>
--	--

1. ☐ Fee Transmittal Form (e.g., PTO/SB/17) (Submit an original, and a duplicate for fee processing)
2. ☒ Specification [Total Pages [36]]
3. ☒ Drawing(s) (35 USC 113) [Total Sheets [1]]
4. ☒ Oath or Declaration [Total Sheets [3]]
  - a. ☐ Newly executed (original or copy)
  - b. ☐ Copy from a prior application (37 CFR 1.63(d))  
(for continuation/divisional with Box 17 completed) [Note Box 5 below]
    - i. ☐ DELETION OF INVENTOR(S) Signed statement attached deleting Inventor(s) named in the prior application, see 37 CFR 1.63(d)(2) and 1.33(b).
5. ☐ Incorporation By Reference (usable if Box 4b is checked) The entire disclosure of the prior application, from which a copy of the oath or declaration is supplied under Box 4b, is considered as being part of the disclosure of the accompanying application and is hereby incorporated by reference therein.
6. ☐ Microfiche Computer Program (Appendix)
7. ☐ Nucleotide and/or Amino Acid Sequence Submission (if applicable, all necessary)
  - a. ☐ Computer Readable Copy b. ☐ Paper Copy (identical to computer copy)
  - c. ☐ Statement verifying identity of above copies

### ACCOMPANYING APPLICATION PARTS

8. ☐ Assignment Papers (cover sheet & document(s))
9. ☐ 37 CFR 3.73 (b) Statement (when there is an assignee) ☐ Power of Attorney
10. ☐ English Translation Document (if applicable)
11. ☐ Information Disclosure Statement (IDS)/PTO-1449 ☐ Copies of IDS Citations
12. ☐ Preliminary Amendment
13. ☒ Return Receipt Postcard (MPEP 503) (Should be specifically itemized)
14. ☐ Small Entity Statement(s) ☐ Statement filed in prior application. Status still proper and desired
15. ☐ Certified Copy of Priority Document(s) (if foreign priority is claimed)
16. ☐ Other: a) Make the references on the parent application(s) of record herein (see MPEP 2001.06(b)).  
b)

17. If a **CONTINUING APPLICATION**, check appropriate box and supply the requisite information:

☐ Continuation ☐ Divisional ☐ Continuation-in-part (CIP) of prior application No: \_\_\_\_\_

### 18. CORRESPONDENCE ADDRESS

☐ Customer Number or Bar Code Label 545 or ☒ Correspondence address below  
(Insert Customer No. or Attach bar code label here)

**Anthony H. Handal** **Reg No.: 26,275**  
**Handal & Morofsky**  
**80 Washington Street** **Telephone: (203) 838-8589**  
**Norwalk, CT 06854 USA** **Fax: (203) 838-8794**

04/21/00  
JCS03 U.S. PTO

09/553810  
04/21/00

HANDAL & MOROFSKY  
Counsellors at Law  
80 Washington Street  
Norwalk, Connecticut 06854



Paul Morofsky 1926-1986

District of Columbia Office  
American Center - Tysons Corner  
Suite 500  
8300 Boone Boulevard  
Vienna, Virginia 22182

Telephone: 203-838-8589  
Facsimile: 203-838-8794  
Telex: 429283 PATSLAW

April 21, 2000

Express Mail Label No. EL465269058US

BOX PATENT APPLICATION  
Assistant Commissioner for Patents  
Washington, D.C. 20231

Re: WILSON et. al.  
New U.S. Non-Provisional Patent Application  
Filed: Herewith Atty. Dkt. No. WILSONLESSAC  
"SPEECH RECOGNITION METHOD"

**NO FEES ARE AUTHORIZED OR DUE TO BE PAID AT THIS TIME**

SIR:

We enclose herewith:

- ☒ Utility Patent Application Transmittal
- ☒ New Patent Application including Specification with Claims (35 pages) and Abstract ( 1 page)
- ☒ One (1 ) Sheet of drawings
- ☒ Declaration
- ☒ Acknowledgment Postcard

Applicant hereby petitions under 37 CFR 1.136 or other applicable rule to have the response period extended the number of months necessary to render the attached communication timely if a petition is required.

Respectfully submitted,

A handwritten signature in black ink, appearing to read "Anthony H. Handal", is written over a horizontal line.

Anthony H. Handal  
Reg. No. 26,275

Inventors:

**Donald WILSON**  
115 Havilands Lane  
White Plains, NY 10605

**Anthony H. HANDAL**  
3 Blue Chip Lane  
Westport CT 06880

**Michael LESSAC**

SPEECH RECOGNITION METHOD

## TECHNICAL FIELD

1 The present invention relates to speech recognition technology of the type  
2 typically embodied in speech recognition software typically implemented on  
3 personal computer systems.

## BACKGROUND

4  
5  
6 In 1964, a group of computer scientists marveled over the new computer being  
7 delivered into the computer center at the Courant Institute for Mathematical  
8 Sciences at New York University. The machine was the latest introduction from  
9 Control Data Corporation, a Model CDC 6600, whose speed and memory  
10 capacity far outstripped the 7.094 K random access memory capacity of the now  
11 humble IBM 7094 that it replaced. In a portent of things to come, they little  
12 suspected that IBM, within months, would obsolete the long heralded CDC 6600  
13 with its IBM 360, a machine which, incredibly, had an unheard-of 360 K of RAM,  
14 all built with discrete components, a conservative move in the face of concerns  
15 about the reliability of the then-new integrated circuit technology. This  
16 impressive machine came to be housed in a room about eighteen feet square,  
17 and surrounded by ten or so air conditioners necessary to keep the system from  
18 overheating and failing. A half dozen tape decks, nearly a meter across and as

1 tall as a man, and several key punch machines, the size of garment outer table  
2 serving machines completed the installation.

3  
4 Thirty-five years later, changes in technology have been remarkable. Tiny  
5 laptop computing devices fly at speeds a thousand times that of those early  
6 powerhouse computers and boast thousands of times the memory. Instead of  
7 huge reels of recording tape, hard disks with capacities on the order of eighteen  
8 GB are found in those same laptop computing devices. These devices, with their  
9 huge memory and computing capabilities, move as freely in the business world  
10 as people, under the arm, in a bag, or on the lap of a businessman flying across  
11 the ocean. No doubt, this technology lies at the foundations of the most  
12 remarkable, reliable and completely unanticipated bull market in the history of  
13 business.

14  
15 Just as certainly, the future holds the promise of similar progress.

16  
17 Notwithstanding the gargantuan magnitude of the progress made in computing  
18 during the last third of the 20th century, the world of computing has been  
19 largely self-contained. The vast majority of all computing tasks involved

1 computers talking to other computers, or otherwise communicating through use  
2 of electrical input signals whose characteristics are substantially absolutely  
3 determined. In this respect, computers are completely unlike the humans which  
4 they serve, humans whose communications vary in almost infinite ways,  
5 regardless of the method of communication, be it voice, writing, or other means.  
6 If computing is to continue to make progress, computers must become  
7 integrated into usual human communications modalities.

8  
9 And, indeed, this is already happening. From a slow start at becoming an  
10 important factor in the marketplace about a decade ago, speech recognition  
11 technology holds just such a promise. A human voice interface to a computer  
12 represents what should be probably the most ideal evolutionarily defined  
13 modality for the human-computer communications interface. While humans  
14 customarily write, gesture and, to a limited extent, use other communications  
15 modes, voice communication remains predominant. This is not surprising,  
16 insofar as speech has evolved in human beings probably for many millions of  
17 years. This is believed to be the case because even relatively primitive forms of  
18 life have fairly highly developed "speech" characteristics. For example, much  
19 work has been done in the study of the use of sounds to communicate various

1 items of information by whales. Likewise, scientists have identified and  
2 cataloged uniform global communications patterns in chimpanzees.

3  
4 In view of the highly natural nature of communication by speech, a direct result  
5 of its having evolved over such a large fraction of the history of the species,  
6 speech communications impose an extremely low level of cognitive overhead on  
7 the brain, thus providing a facile communications interface while allowing the  
8 brain to perform a number of other functions simultaneously. We see this in  
9 everyday life. For example, people engaged in sports activities routinely  
10 combine complex physical tasks, situational analysis, and exchange of  
11 information through speech, sometimes simultaneously transmitting and  
12 receiving audible information, while doing all of these other tasks.

13  
14 Clearly, the mind is well adapted to simultaneously control other tasks while  
15 communicating and receiving audible information in the form of speech. It is  
16 thus no surprise that virtually every culture on earth has devised its own highly  
17 sophisticated audible language.

18  
19 In view of the above, it is thus, easily understood why voice recognition

1 technology has come to be the Holy Grail of computing. While useful work  
2 began to be done with this technology about ten years ago, users only obtained  
3 performance which left much to hope for. Individual quirks, regional  
4 pronunciations, speech defects and impediments, bad habits and the like pepper  
5 virtually everybody's speech to some extent. And this is no small matter. Good  
6 speech recognition requires not only good technology, it also requires  
7 recognizable speech.

8  
9 Toward this end, speech recognition programs generally have an error  
10 correction dialog window which is used to train the system to the features of an  
11 individual user's voice, as will be more fully described below. The motivation  
12 behind such technique is apparent when one considers and analyzes the  
13 schemes typically used in speech recognition systems.  
14

15 Early on, speech recognition was proposed through the use of a series of  
16 bandpass filters. These proposals grew out of the use of spectrographic analysis  
17 for the purpose of speaker identification. More particularly, it was discovered  
18 that if one made a spectral print of a person saying a particular word, wherein  
19 the x-axis represented time and y-axis represented frequency, with the intensity



1 of sound at the various frequencies being displayed in shades of gray or in black  
2 and white, the pattern made by almost every speaker was unique, largely as a  
3 function of physiology, and speakers could be identified by their spectrographic  
4 "prints". Interestingly enough, however, the very diversity which this technique  
5 showed suggested to persons working in the field the likelihood that  
6 commonalities, as opposed to differences, could be used to identify words  
7 regardless of speaker. Hence the proposal for a series of bandpass filters to  
8 generate spectrographs for the purpose of speech recognition.

9  
10 While such an approach was logical given the state of technology in the 1960s,  
11 the problems were also apparent. Obtaining high-quality factors or "Q" in  
12 electrical filters comprising inductors and capacitors is extremely difficult at  
13 audio frequencies.

14  
15 This is due to a number of factors. First of all, obtaining resonance at these  
16 frequencies necessitates the use of large capacitors and inductors. Such  
17 components, in the case of capacitors have substantial resistance leak through.  
18 In the case of inductances, large values of inductance are required, thus  
19 requiring large lengths of wire for the windings and, accordingly, high

1 resistance. The result is that the selectivity of the filters is extremely poor and  
2 the ability to separate different bandpasses is compromised. Finally, the  
3 approach was almost fatally flawed, from a mass-market standpoint, by the fact  
4 that these tuned electrical circuits were very large and mechanically  
5 cumbersome, as well as very expensive.

6  
7 However, in the late 1960's, electrical engineers began to model the action of  
8 electrical circuits in the digital domain. This work was done by determining,  
9 using classical analytic techniques, the mathematical characteristics of the  
10 electrical circuit, and then solving these equations for various electrical inputs.  
11 In the 1970's, it was well understood that the emerging digital technology was  
12 going to be powerful enough to perform a wide variety of computing tasks  
13 previously defaulted to the analog world. Thus, it was inevitable that the  
14 original approaches to voice recognition through the concept of using banks of  
15 tuned circuits would eventually come to be executed in the digital domain.

16  
17 In a typical speech recognition system, an acoustic signal received by a  
18 microphone is input into a voice board which digitizes the signal. The computer  
19 then generates a spectrogram which, for a series of discrete time intervals,

1 records those frequency ranges at which sound exists and the intensity of sound  
2 in each of those frequency ranges. The spectrogram, referred to in the art as a  
3 token, is thus a series of spectrographic displays, one for each of a plurality of  
4 time intervals which together form an audible sound to be recognized. Each  
5 spectrographic display shows the distribution of energy as a function of  
6 frequency during the time interval. In a typical system, sampling rates of 6,000  
7 to 16,000 samples per second are typical, and are used to generate about fifty  
8 spectrum intervals per second for an audible sound to be recognized.

9  
10 In a typical system, quantitative spectral analysis is done for seven frequency  
11 ranges, resulting in eight spectral parameters for each fiftieth of a second, or  
12 spectral sample period. While the idea that a spectral analysis over time can be  
13 a reliable recognition strategy may be counterintuitive given the human  
14 perspective of listening to envelope, tonal variation and inflection, an objective  
15 view of the strategy shows that exactly this information is laid out in an easy to  
16 process spectral analysis matrix.

17  
18 Based on the theoretical underpinnings of the above recognition strategy,  
19 development of a speech recognition system involves the input of vocabulary

1 into the hard drive of a computer in the form of the above described spectral  
2 analysis matrix, with one or more spectral analysis matrices for each word in the  
3 vocabulary of the system. These matrices then serve as word models.

4  
5 In more advanced systems (such as those using so-called "natural" speech, that is  
6 continuous strings of words, the natural tendency of speakers to, on occasion,  
7 blend the end of one word into the beginning of another, and less frequently to  
8 separate words into two parts, sometimes with association of the parts with  
9 different words) models are also developed for these artifacts of the language to  
10 be recognized.

11  
12 Once broken down into a spectral picture over time of frequency energy  
13 distributions, recognition of speech is reduced to comparison of known spectral  
14 pictures for particular sounds to the sound to be recognized, and achieving  
15 recognition through the determination of that model which best matches the  
16 unknown speech sound to be recognized. But this picture, while in principle  
17 correct, is an unrealistic simplification of the problem of speech recognition.

18  
19 After a database of word models has been input into the system, comparison of

1 an audible sound to the models in the database can be used as a reliable means  
2 for speech recognition. However, there are many differences in the speech  
3 patterns of users. For example, different speakers speak at different rates. Thus,  
4 for one speaker, a word they take a certain period of time, while for another  
5 speaker, the word they take a longer period of time. Moreover, different  
6 speakers have voices of different pitch. In addition, speakers may give different  
7 inflections, emphasis, duration and so forth to different syllables of a word in  
8 different ways, depending on the speaker. Even a single speaker will speak in  
9 different ways on different occasions.

10  
11 Accordingly, effective speech recognition requires normalization of spoken  
12 sounds to word and phrase models in the database. In other words, the encoded  
13 received sound or token must be normalized to have a duration equal to that of  
14 the model. This technology is referred to as time aligning, and results in  
15 stretching out or compressing the spoken sound or word to fit it against the  
16 model of the word or sound with the objective of achieving the best match  
17 between the model and the sound input into the system. Of course, it is possible  
18 to leave the sound unchanged and stretch or compress the model.

1 In accordance with existing technology, each of the spectral sample periods for  
2 the sound to be recognized are compared against the corresponding spectral  
3 sample periods of the model which is being rated. The cumulative score for all  
4 of the sample periods in the sound against the model is a quality rating for the  
5 match. In accordance with existing technology, the quality ratings for all the  
6 proposed matches are compared and the proposed match having the highest  
7 quality rating is output to the system, usually in the form of a computer display  
8 of the word or phrase.

9  
10 However, even this relatively complex system fails to achieve adequate quality  
11 in the recognition of human speech. Accordingly, most commercial systems, do  
12 a contextual analysis and also require or strongly recommend a period of  
13 additional training, during which the above matching functions are performed  
14 with respect to a preselected text. During this process, the model is appended to  
15 take into account the individual characteristics of the person training the system.  
16 Finally, during use, an error correction dialog box is used when the user detects  
17 an error, inputs this information into the system and thus causes the word  
18 model to become adapted to the user's speech. This supplemental training of the  
19 system may also be enhanced by inviting the user, during the error correction

1 dialog to speak the word, as well as other words that may be confused with the  
2 word by the system, into the system to further train the recognition engine.

3  
4 As is apparent from the above discussion, the development of speech  
5 recognition systems has centered on assembling a database of sound models  
6 likely to have a high degree of correlation to the speech to be recognized by the  
7 speech recognition engine. Such assembly of the database takes two forms. The  
8 first is the input of global information using one or more speakers to develop a  
9 global database. The second method in which the database is assembled is the  
10 training of the database to a particular user's speech, typically done both during  
11 a training session with preselected text, and on an *ad hoc* basis through use of the  
12 error correction dialog window in the speech recognition program.

#### 13 14 SUMMARY OF THE INVENTION

15 In accordance with the invention, the performance of the speech recognition  
16 software is improved by focusing in on the user, as opposed to the software. In  
17 particular, the invention has its objective improvement of the speech patterns of  
18 persons using the software. The result is enhanced performance, with the bonus  
19 of voice training for the user. Such training is of great importance. For example,

1 salesman, lawyers, store clerks, mothers dealing with children and many others  
2 rely heavily on oral communications skills to accomplish daily objectives.

3 Nevertheless, many individuals possess poor speaking characteristics and take  
4 this handicap with them to the workplace and throughout daily life.

5  
6 Perhaps even more seriously, speech defects, regionalisms, and artifacts  
7 indicative of social standing, ethnic background and level of education often  
8 hold back persons otherwise eminently qualified to advance themselves in life.

9 For this reason, speech, as a subject, has long been a part of the curricula in  
10 many schools, although in recent years education in this area has become, more  
11 and more, relegated to courses of study highly dependent on good speaking  
12 ability, such as radio, television, motion pictures and the theater.

13  
14 Part of the problem here has been the difficulty of finding good voice instructors  
15 and the relatively high cost of the individualized instruction needed for a high  
16 degree of effectiveness in this area. In accordance with the invention, a  
17 specialized but highly effective speech training regimen is provided for  
18 application in the context of speech recognition software for receiving human  
19 language inputs in audio form to a microphone, analyzing the same in a



1 personal computer and outputting alphanumeric documents and navigation  
2 commands for control of the person computer.  
3

4 In accordance with a present invention speech recognition is performed on a  
5 first computing device using a microphone to receive audible sounds input by a  
6 user into a first computing device having a program with a database consisting  
7 of (i) digital representations of known audible sounds and associated  
8 alphanumeric representations of the known audible sounds and (ii) digital  
9 representations of known audible sounds corresponding to mispronunciations  
10 resulting from known classes of mispronounced words and phrases. The  
11 method is performed by receiving the audible sounds in the form of the  
12 electrical output of a microphone. A particular audible sound to be recognized is  
13 converting into a digital representation of the audible sound.  
14

15 The digital representation of the particular audible sound is then compared to  
16 the digital representations of the known audible sounds to determine which of  
17 those known audible sounds is most likely to be the particular audible sound  
18 being compared to the sounds in the database. A speech recognition output  
19 consisting of the alphanumeric representation associated with the audible sound

1 most likely to be the particular audible sound is then produced. An error  
2 indication is then received from the user indicating that there is an error in  
3 recognition. The user also indicates the proper alphanumeric representation of  
4 the particular audible sound. This allows the system to determine whether the  
5 error is a result of a known type or instance of mispronunciation. In response to  
6 a determination of error corresponding to a known type or instance of  
7 mispronunciation, the system presents an interactive training program from the  
8 computer to the user to enable the user to correct such mispronunciation.

9  
10 The presented interactive training program comprises playback of the properly  
11 pronounced sound from a database of recorded sounds corresponding to proper  
12 pronunciations of the mispronunciations resulting from the known classes of  
13 mispronounced words and phrases.

14  
15 In accordance with a preferred embodiment of the invention, the user is given  
16 the option of receiving speech training or training the program to recognize the  
17 user's speech pattern, although this is the choice of the user of the program.

18  
19 In accordance with the invention, the determination of whether the error is a

1 result of a known type or instance of mispronunciation is performed by  
2 comparing the mispronunciation to the digital representations of known audible  
3 sounds corresponding to mispronunciations resulting from known classes of  
4 mispronounced words and phrases using a speech recognition engine.

5  
6 It is anticipated that the inventive method will be implemented by having the  
7 database consisting of (i) digital representations of known audible sounds and  
8 associated alphanumeric representations of the known audible sounds and (ii)  
9 digital representations of known audible sounds corresponding to  
10 mispronunciations resulting from known classes of mispronounced words and  
11 phrases, generated by the steps of speaking and digitizing the known audible  
12 sounds and the known audible sounds corresponding to mispronunciations  
13 resulting from known classes of mispronounced words and phrases. The  
14 database will then be introduced into the computing device of many users after  
15 the generation by speaking and digitizing has been done on another computing  
16 device and transferred together with voice recognition and error correcting  
17 subroutines to the first computing device using CD-ROM or other appropriate  
18 data carrying medium.

1 It is also contemplated that mispronunciations are input into the database by  
2 actual speakers that have such errors as a natural part of this speech pattern.  
3

4 In accordance with the invention, normalization to word, phrase and other  
5 sound models may be achieved by normalizing words or phrases to one of a  
6 plurality of sound durations. This procedure is followed with respect to all the  
7 word that phrase models in the database. When a word is received by the  
8 system, it measures the actual duration, and then normalizes the duration of the  
9 sound to one of the plurality of the preselected normalized sound durations.  
10 This reduces the number of items in the database against which the sound is  
11 compared and rated.  
12

### 13 BRIEF DESCRIPTION OF THE DRAWINGS

14 One way of carrying out the invention is described below in connection with the  
15 figures, it which:  
16

17 Figure 1 is a block diagram illustrating a voice recognition program in  
18 accordance with the method of the present invention;  
19

## 1 DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

2 Referring to Figure 1, the system and method of the present invention may be  
3 understood. In accordance with the inventive method 10, a voice and error  
4 model is generated using subroutine 12. Subroutine 12 comprises a number of  
5 steps which are performed at the site of the software developer, the results of  
6 which are sent for example, in the form of a CD-ROM, other media or via the  
7 Internet, together with the software for executing voice recognition, to a user, as  
8 will be apparent from the description below. In accordance with the present  
9 invention, the inventive speech recognition method may be practiced on  
10 personal computers, as well as on more advanced systems, and even relatively  
11 stripped down lightweight systems, the referred to as subnotebooks and even  
12 smaller systems provided at the same have sound boards for interfacing with  
13 and receiving the output of a microphone. It is noted that quality sound board  
14 electronics is important to good recognition.

## 16 SOFTWARE DEVELOPMENT PHASE

17 At step 14 a database of word models is generated by having speakers speak the  
18 relevant words and phrases into a microphone connected to the sound board of  
19 the personal computer being used to generate the database. In accordance with

1 the preferred embodiment of the invention, speakers who have been trained in  
2 proper speech habits are used to input words, phrases and sounds into the  
3 database at step 14. As the information is generated by the speakers speaking  
4 into microphones, attached to the sound boards in the computer, the  
5 information is digitized, analyzed and stored on the hard drive 16 of the  
6 computer.

7  
8 In accordance with the present invention, relatively common pronunciation  
9 errors are also input into the system at step 18. In this specification the term  
10 "phoneme" is used to mean the smallest sound, perhaps meaningless in itself,  
11 capable of indicating a difference in meaning between two words. The word  
12 "dog" differs from "cog" by virtue of a change of the phoneme "do" pronounced  
13 "daw" and "co" pronounced "cah."

14  
15 Thus, at step 18, the model generating speaker can speak a database of common  
16 phoneme errors into the microphone attached to the sound board of the  
17 computer to result in input of an error database into hard drive 16 of the  
18 computer. However, it is preferred that the phoneme errors are spoken by  
19 persons who in various ways make the pronunciation error as part of their  
20 normal speech patterns.

21  
22 At step 20, the system is enhanced by the introduction into the database

1 contained on hard drive 16 of a plurality of exercise word models, selected for  
2 the purpose of training the speech of a user of the system. The same are input  
3 into the system through the use of a microphone and sound board, in the same  
4 way that the database of the language model was input into the system.

5 Generally, a collection of word and/or phrase models is associated with each  
6 type of phoneme error. This is because if a person makes a speech  
7 pronunciation error of a particular type, it is likely that the same speaker makes  
8 certain other errors which have common characteristics with other  
9 pronunciation errors in the group. For example, a person who mispronounces  
10 the word... to sound like... is also likely to mispronounce the words...

11  
12 Exercise phrase models are input at step 22. These exercise phrase models are  
13 stored by the system in hard drive 16. The exercise word models and the  
14 exercise phrase models input into the system at steps 20 and 22, respectively are  
15 associated in groups having common mispronunciation characteristics. The  
16 same are input into the system through the use of a microphone and sound  
17 board, in the same way that the database of the language model was input into  
18 the system.

19  
20 In addition, in accordance with the present invention, it is recognized that  
21 computer errors may result in misrecognition of a particular error, mistaken  
22 acceptance of a mispronunciation, or mistaken rejection of a proper  
23 pronunciation. Accordingly, during the database generation session during  
24 which properly pronounced exercise word models or exercise phrase models or  
25 input into the system at steps 20 and 22, audio recordings of the same are also

1 stored on hard disk 16, to allow for playback of these proper pronunciations  
2 during use the program by a person performing speech recognition using the  
3 program. This provides for an audible cue to the user and allows the user to  
4 monitor the reliability of the system during the voice recognition and speech  
5 training process of the present invention.  
6

7 In accordance with the invention, it is anticipated that there may be more than  
8 one mispronunciation associated with a particular word or phrase. Accordingly,  
9 at step 24, a plurality of typical mispronunciations are input into the system to  
10 create a database of exercise word error models in hard drive 16. The same are  
11 input into the system through the use of a microphone and sound board, in the  
12 same way that the database of the language model was input into the system.  
13

14 Finally, the database of relatively common mispronunciation errors is completed  
15 at step 26 where the speaker generating that database speaks into the system to  
16 generate a plurality of exercise phrase error models. These error models are also  
17 input into the system through the use of a microphone and stored on hard drive  
18 16.  
19

20 In accordance with a preferred embodiment of the invention, the input of  
21 audible sounds into the system to generate the word error models at step 24 and  
22 the exercise phrase error models at step 26 is done using a speaker or speakers  
23 who have the actual speech error as part of their normal speech patterns. The  
24 same is believed to achieve substantially enhanced recognition of speech errors,  
25 although the same is not believed to be necessary to a functioning system.



1 In accordance with the preferred embodiment of the invention, the models  
2 stored on hard disk 16, and generated as described above may be recorded on a  
3 CD-ROM or other program carrying media, together with a voice recognition  
4 engine, such as that marketed by any one of a number of manufacturers such as  
5 IBM, Dragon Systems, and others. In accordance with a present invention, such  
6 a prior art speech recognition program may be used for both the purpose of  
7 recognizing words, recognizing mispronunciations and phoneme errors,  
8 together with the above described audio recordings of proper pronunciations,  
9 both during speech recognition operation training sessions.

10  
11 In accordance with the invention, such software comprising the speech  
12 recognition engine, editing and training utilities, and database of word models,  
13 phrase models, vocal recordings, and error models may be supplied to the user  
14 for a one time fee and transported over a publicly accessible digital network,  
15 such as the Internet. Alternatively, the software may be made available for  
16 limited use for any period of time with charges associated with each such use, in  
17 which case the software would never be permanently resident on the computer  
18 of a user.

#### 19 20 USER TRAINING PROGRAM

21  
22 When a user desires to use the inventive program, the software containing the  
23 program and the database is loaded into a personal computer and words are  
24 spoken into a microphone coupled to the sound board of the computer, in order  
25 to input the speech into the computer in the manner of a conventional speech

1 recognition program.

2  
3 More particularly, as discussed above, after the system has proceeded through  
4 the performance of steps 14, 18, 20, 22, 24 and 26, and the speech recognition  
5 engine, editing and training utilities added, the system proceeds at step 28 to  
6 receive, through a microphone, speech to be recognized from a user of the  
7 program who has loaded the speech recognition engine, editing and training  
8 utilities, and database of word models, phrase models, vocal recordings, and  
9 error models onto the user's personal computer. In this respect, the operation of  
10 the speech recognition program of the present invention is substantially  
11 identical to other speech recognition programs presently on the market. More  
12 particularly, at step 30, a conventional speech recognition algorithm is applied  
13 to recognize audible sounds as the words which they are meant to represent.

14  
15 The computer then outputs the recognized speech on the screen of the computer  
16 monitor, and the next phrase uttered by the user proceeds at step 30 through the  
17 speech recognition algorithm resulting in that speech also being displayed on  
18 the monitor screen. When the user notices that an error has occurred, he may  
19 use any one of a number of different techniques to bring up an error correction  
20 window at step 32. For example, he may simply double-click on the error, or  
21 highlight the erroneous recognition and hit a key dedicated to presentation of  
22 the error correction window.

23  
24 User correction occurs at step 34. In typical programs, call up of the error  
25 correction window results in the presentation of a screen showing the

1 highlighted word, and suggesting, through the use of a menu, a number of  
2 alternatives which may be selected for double-clicking, in order to correct the  
3 error. If the problem word is not in the menu of alternatives, the user may type  
4 in the problem word or spell it out. After the system has been given the correct  
5 word by any of these means, the same is input into the system.

6  
7 At this point, the call up of the error correction window at step 34 has indicated  
8 to the system that there is an error. While some errors are unrelated to  
9 pronunciation errors, many are. Once the user indicates the error, the system  
10 then proceeds at step 36 to determine whether the error made by the user is  
11 recognized as one of the speech errors recognized by the system. If it is, this  
12 information is determined at step 36. That the nature of the pronunciation error  
13 is then input into the system and logged at step 38. In this matter, the system  
14 keeps track of the number of errors of a particular type for the user by storing  
15 them and tallying them in hard drive 16.

16  
17 In accordance with the present invention, it is contemplated that the speech  
18 training will not be triggered by a single mispronunciation. Instead, it is  
19 contemplated that the repeated instances of a single type of mispronunciation  
20 error will be tallied, and when a threshold of pronunciation errors is reached in  
21 the tally, only then will speech training be proposed by the appearance of the  
22 screen of a prompt window suggesting speech training. The same could take the  
23 form of a window having the words "The system has determined that it is likely  
24 that we can improve your recognition by coaching you. Would you like to  
25 speak to the speech coach?" The screen may also have a headline above the

1 question, such as "The coach wants to talk to you!" The screen will also have a  
2 button bar "OK" and another marked "Cancel", to give the user the opportunity  
3 to click on the "OK" button to start a training session, or to click on the "Cancel"  
4 button to cancel the speech coaching session.

5  
6 It is also noted that other combinations of events may be used to trigger training.  
7 For example, if the particular mispronunciation detected is a very well-defined  
8 one, such as the almost uniform tendency of some speakers to mispronounce the  
9 word "oil" as "earl", the definiteness with which this error has been determined  
10 makes training relatively likely to be necessary, and the threshold can be  
11 lowered to, for example, one instance of that error being detected. In other  
12 cases, or in the general case, one may wish to set the threshold at three, five or  
13 even ten instances of the error before the "The coach wants to talk to you!"  
14 screen is presented to the user of the system.

15  
16 Once a mispronunciation has been detected by the system as a result of the  
17 information input by the user to the user correction screen at step 34, the error  
18 correction algorithm operates in a manner identical to the speech recognition  
19 algorithm at step 30, except that the error correction algorithm checks the  
20 database of common phoneme errors input into the system by the software  
21 developer at step 18 and the exercise word error models and exercise phrase  
22 error models input at steps 24 and 26. In connection with this, it is noted that  
23 the so-called phoneme errors relate to particular sounds consisting of one  
24 syllable or less, while the phrase and word models are somewhat more general,  
25 as described herein.

1 Thus, if, at step 40 the system determines that the threshold number of errors in  
2 that class has not been reached, it sends the system back to step 28, where speech  
3 recognition proceeds. If, on the other hand, a predetermined number of errors of  
4 the same class have been detected by the system and logged at step 38, at step 40  
5 the system is sent to step 42 where the above described "The coach wants to talk  
6 to you!" screen is presented to the user, who is thus given the opportunity to  
7 train his voice.

8  
9 If the user declines the opportunity to train at step 42, he is given the  
10 opportunity to train the database at step 43. If he declines that opportunity also,  
11 the system is returned to step 28, where, again, speech recognition proceeds.

12  
13 However, if he accepts the opportunity to train the database, the system  
14 proceeds to step 45, where the database is trained in the same manner as a  
15 conventional speech recognition processing program.

16  
17 In the other case, at step 42, when the user decides to accept speech training, the  
18 system proceeds to step 44, where the determination is made as to whether the  
19 particular error is an error in the pronunciation of a word or what is referred to  
20 herein as a phrase. By "phrase" in this context, is meant at least parts from two  
21 different words. This may mean two or more words, or the combination of one  
22 or more words and at least a syllable from another word, and most often the end  
23 of one word combined with the beginning of another word, following the  
24 tendency of natural speakers to couple sounds to each other, sometimes  
25 varying their stand-alone pronunciation. If, at step 44 the system determines

1 that the mispronunciation is the mispronunciation of a word, the system is sent  
2 to step 46 where the system retrieves from memory words which have the same  
3 or similar mispronunciation errors.  
4

5 As noted above, these words have been stored in the system, not only in the  
6 form of alphanumeric presentations, but also in high-quality audio format. The  
7 object of the storage of the high-quality audio sound is to provide for audible  
8 playback of the words in the training dialog screen.  
9

10 The words retrieved at step 46 are also presented on-screen in alphanumeric  
11 form to the user and the user is invited to pronounce the word. If the word is  
12 pronounced properly, this is determined at step 48. If there is no error, the  
13 system proceeds to step 50 where the system determines whether there are two  
14 incidences of no error having occurred consecutively. If no error has occurred  
15 twice consecutively, the system is returned to act as a voice recognition system  
16 at step 28. If no error has occurred only once, at step 50 the system is returned to  
17 the training dialog screen at step 46 and the user is invited to pronounce the  
18 same or another word having the same type of mispronunciation to ensure that  
19 the user is facet word correctly. Once the user has pronounced words twice in a  
20 row without errors, the user is returned at step 50 to the voice recognition  
21 function.  
22

23 However, where an error has been detected at step 48, the system proceeds to  
24 step 50 to where an instruction screen telling the user how to make the sound,  
25 with physical instructions on how to move the muscles of the mouth and tongue

1 to achieve the sound is presented to the user.

2  
3 The screen allows for the incorporation of more creative speech training  
4 approaches such as the Lessac method described in *The Use and Training of the*  
5 *Human Voice -A Bio-Dynamic Approach to Vocal Life*, Arthur Lessac, Mayfield  
6 Publishing Co. (1997). In this technique the user is encouraged to use his “inner  
7 harmonic sensing.” This enhances the description of a particular sound by  
8 having the user explore how the sound affects the user’s feelings or encourages  
9 the user to some action.

10  
11 In an illustrative example, the Lessac method teaches the sound of the letter “N”  
12 by not only describing the physical requirements but also instructs the user to  
13 liken the sound to the “N” in violin and to “Play this consonant instrument  
14 tunefully.” This screen also has a button which may be clicked to cause the  
15 system to play back the high-quality audio sound from memory, which was  
16 previously recorded during software development, as described above.

17  
18 The system may also incorporate interactive techniques. This approach presents  
19 the user with a wire frame drawings of a human face depicting, amongst other  
20 information, placement of the tongue, movement of the lips, etc. The user may  
21 interactively move the wire frame drawing to get a view from various angles or  
22 cause the sounds to be made slowly so that the “facial” movements can be  
23 carefully observed.

24  
25 The user is then invited to say the sound again, and at step 54, the user says the

1 word into the microphone which is coupled to the computer, which compares  
2 the word to the database for proper pronunciation at determines whether there  
3 is an error in the pronunciation of the word at step 56.  
4

5 If there is error, the system is sent back to step 46 where, again, the word is  
6 displayed and the user invited to say the word into the machine to determine  
7 whether there is error, with the system testing the output to determine whether  
8 it should proceed to speech recognition at step 28, when the standard of two  
9 consecutive correct pronunciations has been reached. If there is no error at step  
10 56, however, the tally is cleared and the system proceeds to step 28, where  
11 normal speech recognition continues.  
12

13 If, at step 44 the system determines that the mispronunciation is the  
14 mispronunciation of a phrase, the system is sent to step 58 where the system  
15 retrieves from memory phrases which have the same or similar  
16 mispronunciation errors.  
17

18 As noted above, these phrases have been stored in the system, not only in the  
19 form of alphanumeric presentations, but also in high-quality audio format. The  
20 object of the storage of the high-quality audio sound is to provide for audible  
21 playback of the words in the training dialog screen.  
22

23 The words retrieved at step 58 are also presented on-screen and alphanumeric  
24 form to the user at the user is divided to pronounce the word. If the word is  
25 pronounced properly, this is determined at step 60. If there is no error, the



1 system proceeds to step 62 where the system determines whether there are two  
2 incidents of no error having occurred. If no error has occurred twice, the system  
3 is returned to act as a voice recognition system at step 28. If no error has  
4 occurred only once, at step 62 the system is returned to the training dialog  
5 screen to at step 58 and the user is invited to pronounce the same or to the word  
6 having the same type of mispronunciation to ensure that the user is facet word  
7 correctly. Once the user has pronounced words twice in a row without errors,  
8 the user is returned at step 62 to the voice recognition function.  
9

10 However, where an error has been detected at step 60, the system proceeds to  
11 step 62 to where an instruction screen telling the user how to make the sound,  
12 with physical instructions on how to move the muscles of the mouth and tongue  
13 to achieve the sound is presented to the user as well as any other techniques  
14 such as the Lessac method described herein above.  
15  
16

17 This screen also has a button which may be clicked to cause the system to  
18 playback the high-quality audio sound from memory, which was previously  
19 recorded during software development, as described above.  
20

21 The user is then invited to say the sound again, and at step 66, the user says the  
22 phrase into the microphone which is coupled to the computer, which compares  
23 the word to the database for proper pronunciation and determines whether  
24 there is an error in the pronunciation of the word at step 68.  
25

1 If there is error, the system is sent back to step 58 where, again, the word is  
2 displayed and the user invited to say the word into the machine to determine  
3 whether there is error, with the system testing the output to determine whether  
4 it should proceed to speech recognition at step 28, when the standard of two  
5 consecutive correct pronunciations has been reached. If there is no error at step  
6 68, however, the tally is cleared and the system proceeds to step 28, where  
7 normal speech recognition continues, the training session having been  
8 completed.

9  
10 While an illustrative embodiment of the invention has been described together  
11 with several alternatives for various parts of the system, it is, of course,  
12 understood that various modifications will be obvious to those of ordinary skill  
13 in the art. Such modifications are within the spirit and scope of the invention,  
14 which is limited and defined only by the following claims.

## Claims:

1. A method of speech recognition using a microphone to receive audible sounds input by a user into a first computing device having a program with a database consisting of (i) digital representations of known audible sounds and associated alphanumeric representations of said known audible sounds and (ii) digital representations of known audible sounds corresponding to mispronunciations resulting from known classes of mispronounced words and phrases, comprising the steps of:

(a) receiving said audible sounds in the form of the electrical output of said microphone;

(b) converting a particular audible sound into a digital representation of said audible sound;

(c) comparing said digital representation of said particular audible sound to said digital representations of said known audible sounds to determine which said known audible sounds is most likely to be the particular audible sound being compared to the sounds in said database;

(d) outputting as a speech recognition output the alphanumeric representations associated with said audible sound most likely to be said particular audible sound;

(e) receiving an error indication from said user indicating that there is an error in recognition;

(f) receiving from said user an indication of the proper alphanumeric representations of said particular audible sound;

(g) determining whether said error is a result of a known type or instance

of mispronunciation; and

(h) in response to a determination of error corresponding to a known type or instance of mispronunciation, presenting an interactive training program from said computer to said user to enable said user to correct such mispronunciation.

2. A method of speech recognition as in claim 1, wherein said interactive training program comprises playback of the properly pronounced sound from a database of recorded sounds corresponding to proper pronunciations of said mispronunciations resulting from said known classes of mispronounced words and phrases.
3. A method of speech recognition as in claim 2, wherein the user is given the option of receiving speech training or training the program to recognize the user's speech pattern.
4. A method of speech recognition as in claim 3, wherein said determination of whether said error is a result of a known type or instance of mispronunciation is performed by comparing the mispronunciation to said digital representations of known audible sounds corresponding to mispronunciations resulting from known classes of mispronounced words and phrases using a speech recognition engine.
5. A method of speech recognition as in claim 1, wherein the user is given the option of receiving speech training or training the program to recognize the

user's speech pattern.

6. A method of speech recognition as in claim 1, wherein said determination of whether said error is a result of a known type or instance of mispronunciation is performed by comparing the mispronunciation to said digital representations of known audible sounds corresponding to mispronunciations resulting from known classes of mispronounced words and phrases using a speech recognition engine.

7. A method of speech recognition as in claim 1, wherein said database consisting of (i) digital representations of known audible sounds and associated alphanumeric representations of said known audible sounds and (ii) digital representations of known audible sounds corresponding to mispronunciations resulting from known classes of mispronounced words and phrases, is generated by the steps of speaking and digitizing said known audible sounds and said known audible sounds corresponding to mispronunciations resulting from known classes of mispronounced words and phrases.

8. A method of speech recognition as in claim 7, wherein said database has been introduced into said computing device after said generation by speaking and digitizing has been done on another computing device and transferred together with voice recognition and error correcting subroutines to first computing device.

## ABSTRACT OF THE DISCLOSURE

In accordance with a present invention speech recognition is disclosed. It uses a microphone to receive audible sounds input by a user into a first computing device having a program with a database consisting of (i) digital representations of known audible sounds and associated alphanumeric representations of the known audible sounds and (ii) digital representations of known audible sounds corresponding to mispronunciations resulting from known classes of mispronounced words and phrases. The method is performed by receiving the audible sounds in the form of the electrical output of the microphone. A particular audible sound to be recognized is converted into a digital representation of the audible sound. The digital representation of the particular audible sound is then compared to the digital representations of the known audible sounds to determine which of those known audible sounds is most likely to be the particular audible sound being compared to the sounds in the database. A speech recognition output consisting of the alphanumeric representation associated with the audible sound most likely to be the particular audible sound is then produced. An error indication is then received from the user indicating that there is an error in recognition. The user also indicates the proper alphanumeric representation of the particular audible sound. This allows assistant to determine whether the error is a result of a known type or instance of mispronunciation. In response to a determination of error corresponding to a known type or instance of mispronunciation, the system presents an interactive training program from the computer to the user to enable the user to correct such mispronunciation.

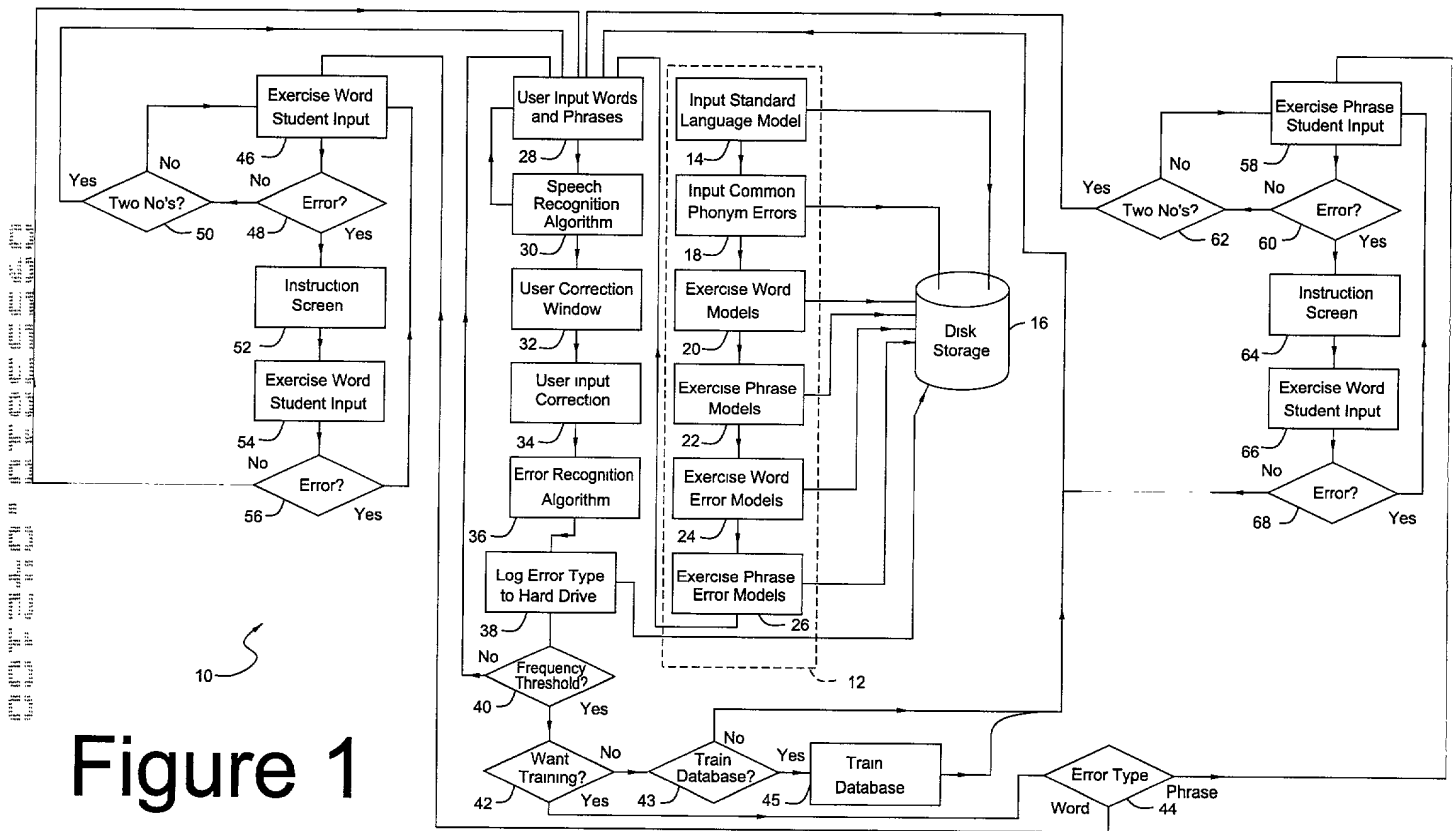


Figure 1

<b>DECLARATION FOR UTILITY OR DESIGN PATENT APPLICATION (37 CFR 1.63)</b>  <input checked="" type="checkbox"/> Declaration Submitted with Initial Filing  OR  <input type="checkbox"/> Declaration Submitted after Initial Filing (surcharge (37 CFR 1.16 (e)) required)	First Named Inventor	WILSON et al.
	COMPLETE IF KNOWN	
	Application Number	Unknown
	Filing Date	Herewith
	Group Art Unit	Unknown
	Examiner Name	Unknown

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name.

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

SPEECH RECOGNITION METHOD

the specification of which

☒ is attached hereto

OR

☐ was filed on \_\_\_\_\_ as United States Application Number or PCT International Application Number and was amended on \_\_\_\_\_ (if applicable).

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment specifically referred to above.

I acknowledge the duty to disclose information which is material to patentability as defined in 37 CFR §1.56.

I hereby claim foreign priority benefits under 35 U.S.C. §119(a)-(d) or §365(b) of any foreign application(s) for patent or inventor's certificate, or §365(a) of any PCT International application which designated at least one country other than the United States of America, listed below and have also identified below, by checking the box, any foreign application for patent or inventor's certificate, or PCT International application having a filing date before that of the application on which priority is claimed.

Prior Foreign Application Number(s)	Country	Foreign Filing Date (mm/dd/yyyy)	Priority Not Claimed	Certified Copy Attached	
			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

I hereby claim the benefit under 35 U.S.C. §119(e) of any United States provisional application(s) listed below.

Application Number(s)	Filing Date (mm/dd/yyyy)



I hereby claim the benefit under 35 U.S.C. §120 of any United States application(s), or 365(c) of any PCT International application designating the United States of America, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT International application in the manner provided by the first paragraph of 35 U.S.C. §112, I acknowledge the duty to disclose information which is material to patentability as defined in 37 CFR §1.56 which became available between the filing date of the prior application and the national or PCT International filing date of this application.

<i>U.S. Parent Application or PCT Parent Number</i>	<i>Parent Filing Date (MM/DD/YYYY)</i>	<i>Parent Patent Number (if applicable)</i>

As a named inventor, I hereby appoint **Anthony H. Handal, Registration No. 26,275,**  
and all other registered attorneys listed under **Customer No. 00545**  
as my attorneys to prosecute this application and to transact all business in the Patent and Trademark Office  
connected therewith.

Address all telephone calls to Anthony H. Handal at (203) 838-8589

Address all correspondence to: Anthony H. Handal  
Handal & Morofsky  
80 Washington Street  
Norwalk, Connecticut 06854  
Facsimile (203) 838-8794

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

NAME OF SOLE OR FIRST INVENTOR:				<input type="checkbox"/> A petition has been filed for this unsigned inventor	
<i>Given Names (at least one name in full plus other names or initials)</i>			<i>Family Name or Surname</i>		
H. Donald			WILSON		
<i>Inventor's Signature:</i>				<i>Date:</i>	
<i>Residence City</i>	White Plains	<i>State</i>	NY	<i>Country</i>	U.S.A
<i>Post Office Address.</i> 115 Havilands Lane					
<i>Post Office Address:</i>					
<i>City</i>	White Plains	<i>State</i>	NY	<i>Zip</i>	10605
				<i>Country</i>	U.S.A

☒ Additional inventors are being named on the separate sheet attached hereto.

<b>NAME OF ADDITIONAL INVENTOR:</b>		<input type="checkbox"/> A petition has been filed for this unsigned inventor	
<i>Given Names (at least one name in full plus other names or initials)</i>		<i>Family Name or Surname</i>	
Anthony H.		HANDAL	
<i>Inventor's Signature:</i>		<i>Date:</i>	
<i>Residence: City</i>	Westport	<i>State</i>	CT
		<i>Country</i>	U.S.A
<i>Post Office Address:</i> 3 Bluechip Lane			
<i>Post Office Address:</i>			
<i>City</i>	Westport	<i>State</i>	CT
		<i>Zip</i>	06880
		<i>Country</i>	U.S.A

<b>NAME ADDITIONAL INVENTOR:</b>		<input type="checkbox"/> A petition has been filed for this unsigned inventor	
<i>Given Names (at least one name in full plus other names or initials)</i>		<i>Family Name or Surname</i>	
Michael		LESSAC	
<i>Inventor's Signature:</i>		<i>Date:</i>	
<i>Residence: City</i>		<i>State</i>	
		<i>Country</i>	
<i>Post Office Address:</i>			
<i>Post Office Address:</i>			
<i>City</i>		<i>State</i>	
		<i>Zip</i>	
		<i>Country</i>	